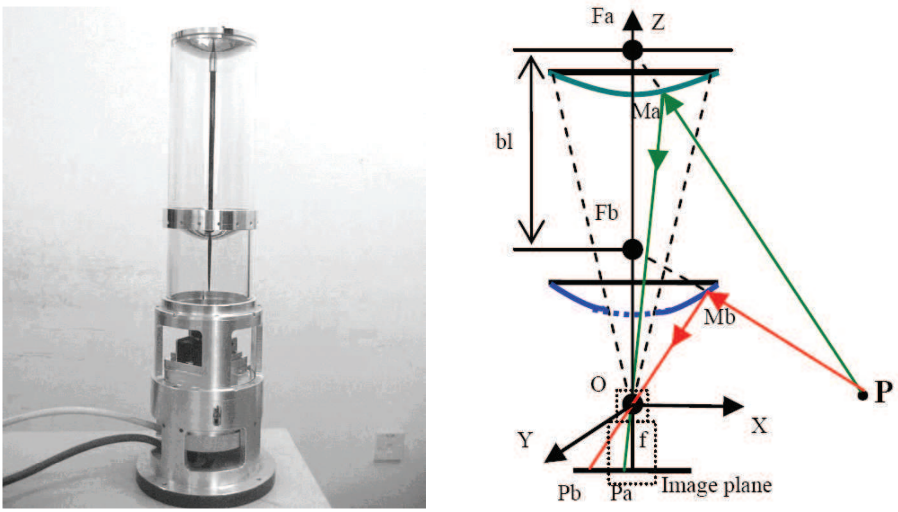
Stereo Matching and 3D Reconstruction via an Omnidirectional Stereo Sensor[1]

Lei He, Chuanjiang Luo, Feng Zhu and Yingming Hao
Shenyang Institute of Automation, Chinese Academy of Sciences
P.R. China

1. Introduction

A catadioptric vision system using diverse mirrors has been a popular means to get panoramic images(K.Nayar, 1997) which contains a full horizontal field of view (FOV). This wide view is ideal for three-dimensional vision tasks such as motion estimation, localization, obstacle detection and mobile robots navigation. Omnidirectional stereo is a suitable sensing method for such tasks because it can acquire images and ranges of surrounding areas simultaneously. For omnidiretional stereo vision, an obvious method is to use two (or more) cameras instead of each conventional camera (K.Tan et al., 2004; J.Gluckman et al., 1998; H.Koyasu et al. 2002; A.Jagmohan et al. 2004). Such two-camera (or more-camera) stereo systems are relatively costly and complicated compared to single camera stereo systems. Omnidirectional stereo based on a double-lobed mirror and a single camera was developed (M.F.D. Southwell et al. 1996; T.L. Conroy & J.B. Moore, 1999; E. L. L. Cabral, et al. 2004; Sooyeong Yi & Narendra Ahuja, 1996) . A double lobed mirror is a coaxial mirror pair, where the centers of both mirrors are collinear with the camera axis, and the mirrors have a profile radially symmetric around this axis. This arrangement has the advantage to produce two panoramic views of the scene in a single image. But the disadvantage of this method is the relatively small baseline it provides. Since the two mirrors are so close together, the effective baseline for stereo calculation is quite small. We have developed a novel omnidirectional stereo vision optical device (OSVOD) based on a common perspective camera coupled with two hyperbolic mirrors, which are separately fixed inside a glass cylinder. As the separation between the two mirrors provides much enlarged baseline, in our system, the baseline length is about 200mm, the precision has improved correspondingly (Fig. 1).

[1] This work is supported by National Science Foundation of P.R. China (60575024).



a b

Figure 1. a: The appearance of the stereo vision system. b: The configuration of the system

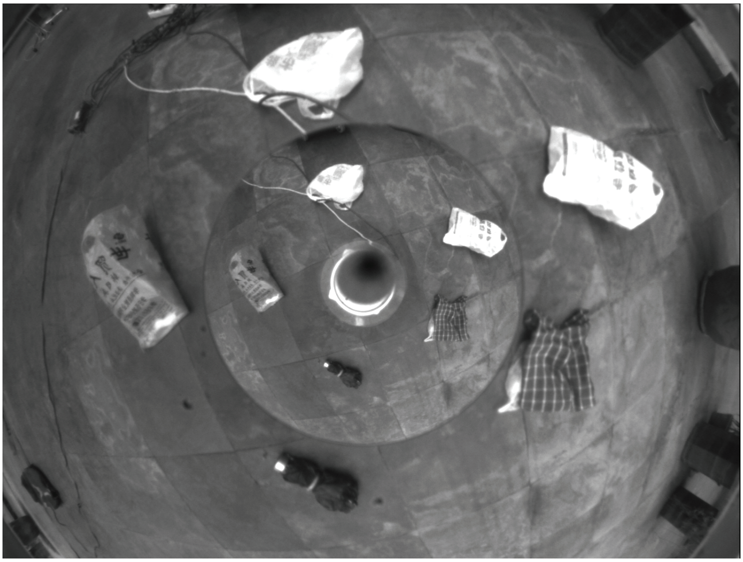


Figure 2. Real indoor scene captured for depth map reconstruction

The coaxial configuration of the camera and the two hyperbolic mirrors makes the epipolar line radially collinear, which makes the system free of the search process for complex epipolar curve in stereo matching (Fig. 2). The OSVOD is mounted on top of mobile robot looking downwards, at a height of approximately 0.75 meters above the ground plane. We also notice that G. Jang et al. (G. Jang, S. Kim & I. Kweon, 2005) proposed a wide-baseline catadioptric stereo system, in which the system design is similar to us but the system design and mirror types are different. Furthermore, for a mobile robot, our system is designed to obtain omnidirectional images that are different from above mentioned stereo systems. Note that our images are taken from a special angle of view to obtain the planform of scene, which aims to reconstruct the depth map that denotes the heigh information (vertical depth) but not the horizontal distance information in common stereo systems. Albeit the calculative precision of triangulation was improved theoretically due to the wider baseline, more complex and difficult stereo matching problem should be brought on because of the wider disparity space that exists and more serious image distortion. A major aim of this paper is to propose an integrated framework, which mainly focuses on stereo matching to enhance the performance for depth map regeneration. Since our primary goal is to propose a precise and suitable algorithm for stereo matching to satisfy the reliability requirement via an omnidirectional stereo vision system, we chiefly review previous stereo matching methods as follows. State of the art algorithms for dense stereo matching can be divided into two categories. One category is local methods, in which some kind of similarity measure over an area is calculated (Devernay & F. Faugeras, 1994). They work well in relatively textured areas in a very fast speed, while they cannot gain correct disparity map in textureless areas and areas with repetitive textures, which is an unavoidable problem in most situations. In (Sara, R., 2002) a method of finding the largest unambiguous component has been proposed, but the density of the disparity map varies greatly depend on the discriminability of the similarity measure in a given situation. The other one is global methods which are generally energy minimization approaches, these methods make explicit smoothness assumptions and try to find a global optimized solution of a predefined energy function that take into account both the matching similarities and smoothness assumptions. Most recent high-perfomance algorithms belong to energy minimization approaches (Pedro F. Felzenszwalb & Daniel P. Huttenlocher, 2006; Y. Boykov et al., 2001; V. Kolmogorov & R. Zabih, 2001; Yedidia, J.S. et al., 2000) due to powerful new optimization algorithms such as graph cuts and loopy belief propagation. The results, especially in stereo, have been dramatic, according to the widely-used Middlebury stereo benchmarks (D. Scharstein & R. Szeliski, 2002), almost all the top-performing stereo methods rely on graph cuts or LBP. Moreover, these methods give substantially more accurate results than what were previously possible. Although numerous methods exist for stereo matching, as to our knowledge, most algorithms are presented and implemented using standard images, there are few algorithms specifically designed for single camera omnidirectional stereo. Sven Fleck et al. (Sven Fleck et al., 2005) completely use a common graph cut method to acquire a 3D model using a mobile robot that is equipped with a laser scanner and a panoramic camera. Other work in point based omnidirectional reconstruction on mobile robotics can be found in (R. Bunschoten & B. Kr¨ose, 2003), however this is not based on graph cuts that we are concerned in this paper. Considering the peculiarities of omnidirectional images, we adapt improved graph cut method, in which a new energy model is introduced for more general priors corresponding to more reasonable piecewise smoothness assumption since the well-known swap move algorithm can be applied to a wider class of energy functions (Y. Boykov et al., 2001). The proposed energy function is different from previous any other ones since the smooth item is based on three variables whereas others only consist of two variables. We also show the necessary modification to handle panoramic images, including deformed matching template, adaptable template scale to elaborate the date term. In the rest of the paper, we first necessarily introduce a full model of calibration in the system. In section 3, our improved optimization model is presented. We generalize our completed omnidirectional stereo matching framework in section 4. In section 5, experiments and their results are given. Finally, we conclude the paper.

2. Calibrating the System

2.1 Principle of Our Vision System

The system we have developed (Su & Zhu, 2005) is based on a common perspective camera coupled with two hyperbolic mirrors, which are separately fixed inside a glass cylinder (Fig.1a). The two hyperbolic mirrors share one focus which coincides with the camera center. A hole in the below mirror permits imaging via the mirror above. As the separation between the two mirrors provides much enlarged baseline, the precision of the system has been improved correspondingly. The coaxial configuration of the camera and the two hyperbolic mirrors makes the epipolar line radially collinear, thus making the system free of the search process for complex epiploar curve in stereo matching (Fig. 3).

To describe the triangulation for computing 3D coordinates of space points, we define the focal point O as the origin of our reference frame, z-axis parallel to the optical axis pointing above. Then mirrors can be represented as:

$$\frac{(z-c_i)^2}{a^2}-\frac{(x^2+y^2)}{b^2}=1, \quad (i=1,2) \tag{1}$$

Only the incident rays pointing to the focus $F_a(0,0,2c_a)$, $F_b(0,0,2c_b)$ will be reflected by the mirrors to pass through the focal point of the camera. The incident ray passing the space point $P(x,y,z)$ reaches the mirrors at points M_a and M_b, being projected onto the image at points $P_a(u_a,v_a,-f)$ and $P_b(u_b,v_b,-f)$ respectively. As P_a and P_b are known, M_a and M_b can be represented by:

$$\frac{x_{M_i}}{u_i}=\frac{y_{M_i}}{v_i}=\frac{z_{M_i}}{-f}, \quad (i=1,2) \tag{2}$$

Since point M_a and M_b are on the mirrors, they satisfy the equation of the mirrors. Their coordinates can be solved from equation group (1) and (2). Then the equation of rays F_aP and F_bP are:

$$\frac{x_p}{x_i}=\frac{y_p}{y_i}=\frac{z_p-2c_i}{z_i-2c_i}, \quad (i=1,2) \tag{3}$$

We can finally figure out coordinate of the space point P by solving the equation (3).

2.2 Overview of Omnidirectional Camera Calibration

In using the omnidirectional stereo vision system, its calibration is important, as in the case of conventional stereo systems (Luong & Faugeras, 1996; Zhang & Faugeras, 1997). We present a full model of the imaging process, which includes the rotation and translation between the camera and mirror, and an algorithm to determine this relative position from observations of known points in a single image.

There have been many works on the calibration of omnidirectional cameras. Some of them are for estimating intrinsic parameters (Ying & Hu, 2004; Geyer & Daniilidis, 1999; Geyer Daniilidis, 2000; Kang, 2000). In (Geyer & Daniilidis, 1999; Geyer Daniilidis, 2000), Geyer & Daniilidis presented a geometric method using two or more sets of parallel lines in one image to determine the camera aspect ratio, a scale factor that is the product of the camera and mirror focal lengths, and the principal point. Kang (Kang, 2000) describes two methods. The first recovers the image center and mirror parabolic parameter from the image of the mirror's circular boundary in one image; of course, this method requires that the mirror's boundary be visible in the image. The second method uses minimization to recover skew in addition to Geyer's parameters. In this method the image measurements are point correspondences in multiple image pairs. Miousik & Pajdla developed methods of calibrating both intrinsic and extrinsic parameters (Miousik & Pajdla, 2003a; Miousik & Pajdla, 2003b). In (Geyer & Daniilidis, 2003), Geyer & Daniilidis developed a method for rectifying omnidirectional image pairs, generating a rectified pair of normal perspective images.

Because the advantages of single viewpoint cameras are only achieved if the mirror axis is aligned with the camera axis, these methods mentioned above all assume that these axes are parallel rather than determining the relative rotation between the mirror and camera. A more complete calibration procedure for a catadioptric camera which estimates the intrinsic camera parameters and the pose of the mirror related to the camera appeared at (Fabrizio et al., 2002), the author used the images of two known radius circles at two different planes in an omnidirectional camera structure to calibrate the intrinsic camera parameters and the camera pose with respect to the mirror. But this proposed technique cannot be easily generalized to all kinds of catadioptric sensors for it requires the two circles be visible on the mirror. Meanwhile, this technique calibrated the intrinsic parameters combined to extrinsic parameters, so there are eleven parameters (five intrinsic parameters and six extrinsic parameters) need to be determined. As the model of projection is nonlinear the computation of the system is so complex that the parameters cannot be determined with good precision.

Our calibration is performed within a general minimization framework, and easily accommodates any combination of mirror and camera. For single viewpoint combinations, the advantages of the single viewpoint can be exploited only if the camera and mirror are assumed to be properly aligned. So for these combinations, the simpler single viewpoint projection model, rather than the full model described here, should be adopted only if the misalignment between the mirror and camera is sufficiently small. In this case, the calibration algorithm that we describe is useful as a software verification of the alignment accuracy.

Our projection model and calibration algorithm separate the conventional camera intrinsics (e.g., focal length, principal point) from the relative position between the mirrors and the camera (i.e., the camera-to-mirrors coordinate transformation) to reduce computational complexity and improve the calibration precision. The conventional camera intrinsics can be determined using any existing method. For the experiments described here, we have used the method implemented in http://www.vision.caltech.edu/bouguetj/calib_doc/. Once the camera intrinsics are known, the camera-to-mirrors transformation can be determined by obtaining an image of calibration targets whose three-dimensional positions are known, and then minimizing the difference between coordinates of the targets and the locations calculated from the targets' images through the projection model. Fig. 3 shows one example of calibration image used in our experiments. The locations of the three dimensional points have been surveyed with an accuracy of about one millimeter. If the inaccuracy of image point due to discrete distribution of pixels is taken into account, the total measuring error is about five millimeters.

2.3 Projection Model

Fig. 3 depicts the full imaging model of a perspective camera with two hyperbolic mirrors. There are three essentially coordinate systems.

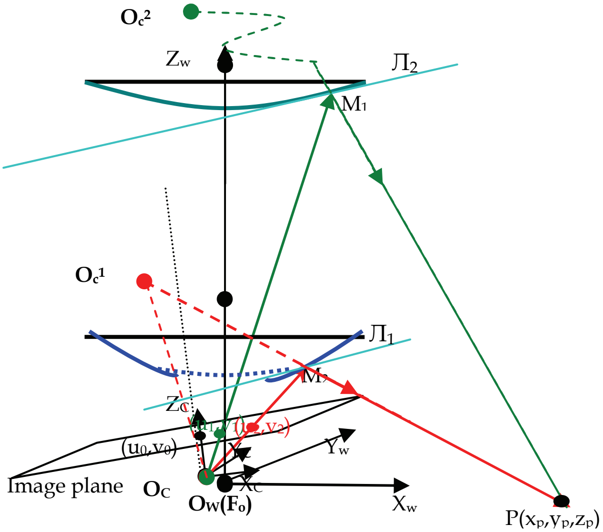


Figure 3. The projection model of the omnidirectional stereo vision system. There are transformations between the camera coordinate system and the mirror (or world) coordinate system

1. The camera coordinate system centered at the camera center O_c, the optical axis is aligned with the z-axis of the camera coordinate system;
2. The mirror system centered at common foci of the hyperbolic mirrors F_o, the mirrors axes is aligned with the z-axis of the mirror coordinate system (We assume that the axes of the mirrors are aligned well, and the common foci are coincident, from the mirrors manufacturing sheet we know it is reasonable);
3. The world system centered at O_w. The omnidirectional stereo vision system was placed on a plane desk. As both the base of vision system and desk surface are plane, the axis of the mirror is perpendicular to the base of the system and the surface of the desk feckly. We make the mirror system coincide with the world system to simplify the model and computations.

So the equations of hyperboloid of two sheets in the system centered at O_w are the same as equation (1). For a known world point $P(x_w,y_w,z_w)$ in the world (or mirror) coordinate system whose projected points in the image plane are also known, $q_1(u_1,v_1)$ and $q_2(u_2,v_2)$ are respectively projected by the upper mirror and bellow mirror. Then we get their coordinates in the camera coordinate system:

$$\begin{bmatrix} x_i^c \\ y_i^c \\ z_i^c \end{bmatrix} = \begin{bmatrix} (u_i-u_0)k_u \\ (v_0-v_i)k_v \\ f \end{bmatrix}, \quad (i=1,2) \tag{4}$$

Where f is the focal length; k_u and k_v are the pixel scale factors; u_0 and v_0 are the coordinates of the principal point, where the optical axis intersects the projection plane. They are intrinsic parameters of the perspective camera.

So the image points $P_c\left(x_i^c, y_i^c, z_i^c\right)$ of the camera coordinate system can be expressed relative to the mirror coordinate system as:

$$\begin{bmatrix} x_i^m \\ y_i^m \\ z_i^m \end{bmatrix} = R \begin{bmatrix} x_i^c \\ y_i^c \\ z_i^c \end{bmatrix} + t, \quad (i=1,2) \tag{5}$$

Where R is a 3×3 rotation matrix with three rotation angles around the x-axis (pitch α), y-axis (yaw β) and z-axis (title χ) of the mirror coordinate system respectively; $t=[t_x,t_y,t_z]$ is the translation vector. So the origin $O_c=[0,0,0]^T$ of the camera coordinate system can be expressed in the world coordinate system $O_m=[t_x,t_y,t_z]^T$, so the equations of lines O_cM_1 and $f=\{f_p|p\in P\}$ which intersect with the upper mirror and bellow mirror respectively at points M_1 and M_2, can be determined by solving simultaneous equations of the line O_cM_1 or O_cM_2 and the hyperboloid. Once the coordinates of the point M_1 and M_2 have been worked out, we can write out the equations of the tangent plane л1 and л2 which passes the upper and the bellow mirror at point M_1 and M_2 respectively. Then the symmetric points O_c^1 and O_c^2 of the origin of the camera coordinate system O_c relative to tangent plane л1 and л2 in the world coordinate system can be solved from the following simultaneous equations:

$$\begin{cases} \dfrac{x_{o_c}{}^i - tx}{a_i{}^2 x_{M_i}} = \dfrac{y_{o_c}{}^i - ty}{a_i{}^2 y_{M_i}} = \dfrac{z_{o_c}{}^i - tz}{-b_i{}^2 z_{M_i} + b_i{}^2 c_i} \\ a_i{}^2 x_{M_i}(tx + x_{o_c}{}^i) + a_i{}^2 y_{M_i}(ty + y_{o_c}{}^i) - (-b_i{}^2 z_{M_i} + b_i{}^2 c_i)(tz + z_{o_c}{}^i), \\ +2[-a_i{}^2 x_{M_i}{}^2 - a_i{}^2 y_{M_i}{}^2 - z_{M_i}(-b_i{}^2 z_{M_i} + b_i{}^2 c_i)] = 0 \end{cases} \quad (i=1,2) \tag{6}$$

Hitherto the incident ray $O_c{}^1M_2$ and $O_c{}^2M_1$ can be written out to determine the world point $P(x_W,y_W,z_W)$. Generally, the two lines are non-co-plane due to various parameter errors and measuring errors, we solve out the midpoint $G=(\hat{x}_W,\hat{y}_W,\hat{z}_W)^T$ of the common perpendicular of the two lines by

$$\left.\begin{array}{l}\begin{cases}[\overrightarrow{O_c{}^1M_2}\times(\overrightarrow{O_c{}^1M_2}\times\overrightarrow{O_c{}^2M_1})]\bullet\overrightarrow{G_1M_2}=0\\ \overrightarrow{G_1M_1}=t\overrightarrow{G_1O_c{}^2}\end{cases}\Rightarrow\overrightarrow{OG_1}\\ \begin{cases}[\overrightarrow{O_c{}^2M_1}\times(\overrightarrow{O_c{}^1M_2}\times\overrightarrow{O_c{}^2M_1})]\bullet\overrightarrow{G_2M_1}=0\\ \overrightarrow{G_2M_2}=t\overrightarrow{G_2O_c{}^1}\end{cases}\Rightarrow\overrightarrow{OG_2}\end{array}\right.\quad \overrightarrow{OG}=(\overrightarrow{OG_1}+\overrightarrow{OG_2})/2 \tag{7}$$

From all of them above, we finally come to the total expression to figure out the world point $G=(\hat{x}_w,\hat{y}_w,\hat{z}_w)^T$ from two image points respectively projected by the upper mirror and bellow mirror and six camera pose parameters left to be determined.

$$G\left(\alpha,\beta,\chi,t_x,t_y,t_z,u_1,v_1,u_2,v_2\right)=\begin{bmatrix}\hat{x}_W\\ \hat{y}_W\\ \hat{z}_W\end{bmatrix} \tag{8}$$

Equation (8) is a very complex nonlinear equation with high power and six unknown parameters to determine. The artificial neural network trained with sets of image points of the calibration targets is used to estimate the camera-to-mirror transformation.

Taking advantage of the ANN capability, which adjusts the initial input camera-to-mirror transformations step by step to minimize the error function, the real transformations parameters of the camera-to-mirror can be identified precisely.

2.4 Error Function

Considering the world points with known coordinates, placed onto a calibration pattern, at the same time, their coordinates can be calculated using the equation (8) from back-projection of their image points. The difference between the positions of the real world coordinates and the calculated coordinates is the calibration error of the model. Minimizing the above error by means of an iterative algorithm such as Levenberg-Marquardt BP algorithm, the camera-to-mirror transformation is calibrated. The initial values for such algorithm are of consequence. In our system, we could assume the transformation between cameras and mirrors is quite small, as the calculation error without considering the camera-to-mirror transformation is not significant thus using R=I and T=0 as the initial values is a reasonable choice.

We minimize the following squared error ε^2:

$$\varepsilon^2=\sum_{i=1}^{n}\left\|P_i-G_i\left(\alpha,\beta,\chi,t_x,t_y,t_z,u_1^i,v_1^i,u_2^i,v_2^i\right)\right\|^2 \tag{9}$$

Where n is the number of the calibration points.

Because $G_i\left(\alpha,\beta,\chi,t_x,t_y,t_z,u_1^i,v_1^i,u_2^i,v_2^i\right)$ depends on the camera-to-mirror transformation, (9) is optimized with respect to the six camera-to-mirror parameters.

2.5 Calibration Result

The calibration was performed using a set of 81 points equally distributed on a desk with different heights from 0 to 122mm around the vision system.

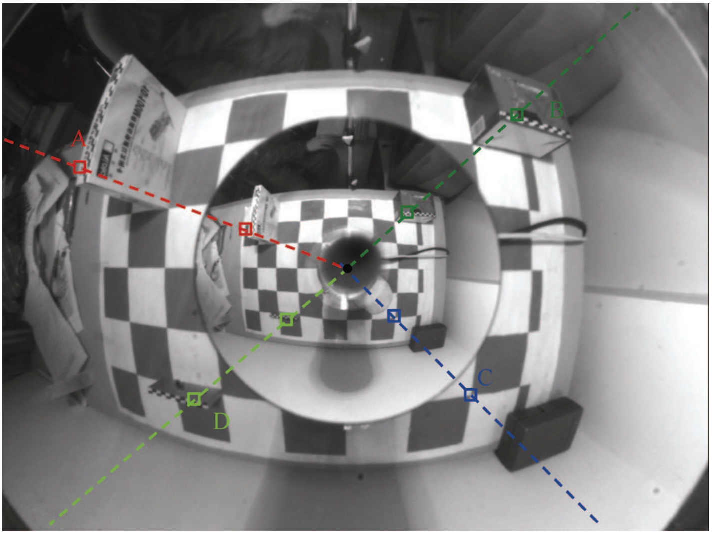


Figure 4. A calibration image used in our experiments. The coaxial configuration of the camera and the two hyperbolic mirrors makes the epipolar line radially collinear, which makes the system free of the search process for complex epipolar curve in stereo matching

The calibration results with real data are listed in Table 1.

	α	β	χ	t_x	t_y	t_z
value	-0.9539°	0.1366°	0.1436°	-0.0553mm	-0.1993mm	1.8717mm

Table 1. Calibration result with real data

The calibration error was estimated using a new set of 40 untrained points, the average square error of the set points is 34.24mm without considering the camera-to-mirror transformation. Then we calculate the error with the transformation values listed in Table 1, the average square error decrease to 12.57mm.

3. Improved Graph-cut Model

In this section, we first briefly introduce the prior work on graph cuts, after that our improved optimization model and corresponding algorithm are presented. Note that our work has been done about graph cuts mainly based on related excellent work in [14], [15], [20] and [21], in our paper, most general depiction is based on above papers.

3.1 Two-variable Smooth Model

In terms of the energy minimization, many early vision problems, which of course contain stereo matching, can be formulated in following energy model:

$$E(f)=E_{data}(f)+E_{smooth}(f). \tag{10}$$

In Eq. (10), the data term $E_{data}(f)$ represents some extent similarities between f and the observed data and typically,

$$E_{data}(f)=\sum_{p\in P} D_p(f_p). \tag{11}$$

Where, P denotes the set of image pixels that need to be assigned labels. The label assigned to pixel $p\in P$ is denoted by f_p, and f is the set of all assignments: $f=\{f_p \mid p\in P\}$. D_p measures how well the label f_p fits the pixel p in the observed image. The smooth term $E_{smooth}(f)$ usually represents the smoothness of f ,which is a critical issue and lots of functions have been proposed, typically it can be expressed as follows,

$$E_{smooth}(f)=\sum_{(p,q)\in N} V_{pq}(f_p,f_q). \tag{12}$$

Where N represents the set of neighboring pixel pairs, in this case, Vpq represents the smoothness of pixels p and q that are respectively denoted by fp and fq, thereby *Esmooth* reflects the smoothness of all the neighboring pixels because our prior knowledge tells us that the surfaces of objects invariably keep relatively smooth except for some discontinuity area. We call the smoothness model that based on this function two-variable smoothness model because E_{smooth} refers to two variables. Considering above description, the following energy function is commonly minimized in computer vision and graphical fields such as image restoration, image segmentation and stereo matching:

$$E(f)=\sum_{p\in P} D_p(f_p)+\sum_{(p,q)\in N} V_{pq}(f_p,f_q). \tag{13}$$

Currently, there are two typical two-variable smoothness models that specify smooth assumption that exist. One is Potts model, in which, the smooth term is depicted by following expression,

$$V_{pq}(f_p,f_q)=C_{pq}\cdot\min(1,\left|f_p-f_q\right|) \tag{14}$$

where once any difference exists in label pairs, the penalty to this difference via smooth energy is bound to be the same amount, theoretically, the expectation of labels f should be constant since only the smoothness assumption is considered, however, the presence of the data term usually results in piecewise constant in fact. The other model is truncated convex priors,

$$V_{pq}(f_p,f_q)=C_{pq}\cdot\min(T,g(f_p-f_q)). \tag{15}$$

Generally, $g(x)$ is convex and symmetric, for linear truncated function, $g(x)=|x|$ and when $g(x)=x^2$, it is truncated quadratic function, this smoothness energy function tend to give a piecewise constant assumption theoretically on smoothness term, albeit practically the piecewise smooth results may come out owing to the data term. In this case, label set f is expected to be consisted of several pieces and in each piece the difference between the adjacent pixels in f is just a little, in other words, the neighboring pixels tend to vary smoothly. The truncation constant T is crucial to limit the penalty on smooth, otherwise, the penalty on neighboring pixel fp and fq might be unboundedly large, this is somewhat ridiculous since discontinuity always exist in most images and it may lead to oversmoothed label set f.

3.2 Three-variable Smooth Model

Our three-variable smoothness model is proposed to solve the problem that exists in two-variable smoothness one. Note that only two variables, which are usually labels of neighboring pixels, are considered in smoothness assumption. Apparently, only the zero order and first order derivatives can be obtained via only two adjacent pixels. Nevertheless, sometimes they are not enough to specify the smoothness of a surface although expecting to minimize the difference of neighboring pixels (lesser first order derivative of f) invariably gives a reasonable constraint to make the whole surface that denoted by f varies relatively smoothly. This constraint is reasonable customarily mainly because the less difference between neighboring pixels likely means the more smoothly f varies. And also the data term, which accounts for biggish proportion in the whole energy function, always tend to enshroud the imprecise smoothness model. However, the smoothness term needs to be more precise to obtain reliable results when the disparity space is large, which means the reconstructed scene surface needs to be more elaborated and when the data term can not contribute to the energy function well in some conditions such as weak textured and textureless area. Unfortunately, these two take place in our omnidirectional images. To represent the smoothness of f better, we propose a three-variable smoothness model in which Esmooth typically can be expressed in following form:

$$E_{smooth}(f)=\sum_{(p,q,r)\in N} V_{pqr}(f_p,f_q,f_r). \tag{16}$$

In Eq. (16), neighboring pixels *p, q* and *r* are series-wound orderly, *Esmooth* contains three variables so that it can represent the smoothness of *f* more appropriately than two-variable model does. Now we give the concrete expression:

$$V_{pq}(f_p,f_q,f_r)=C_{pqr}\cdot\min(T,g(\left|f_p+f_r-2f_q\right|+\left|f_p-f_r\right|)), \tag{17}$$

where g(x) can be defined as the same in Eq. (16), |fp+fr-fq| represents the second derivative of label fr. In this case, the labels between the neighboring pixels tend to vary consistently and that also means less curvature which can represent the smoothness of surface better. Without doubt, to vary steadily is also important, thereby, |fp-fr|, which denotes the offset between first pixel p and last pixel r, is added in Vpqr. Obviously, the labels are expected to vary both smoothly and consistently in three-variable smoothness model while in two-variable smoothness model, the labels are only emphasized on varying smoothly.

3.3 Graph cuts for 3-variable smooth model

As we know, it is NP-hard to optimize the energy functions in Eq. (14) and Eq. (15). To solve this problem, Boykov et al. (Y. Boykov et al., 2001) developed the expansion and swap algorithm in which when Vpq is Potts and truncated linear or quadratic, the energy function can be optimized approximately. Now we use swap algorithm to optimize our 3-variable smoothness model.

Before constructing graph for 3-variable smoothness model, in the first place, we necessarily prove this energy function is graph-representable which is equal to proving Esmooth(f) is regular. According to (V. Kolmogorov & R. Zabih, 2004), we only need to confirm all the projections of Vpqr of two variables are regular. To complete the proof of this conclusion, following three inequalities should be proved, for simplicity, we use V to represent Vpqr.

$$V(\alpha,\alpha,f_r)+V(\beta,\beta,f_r)\le V(\alpha,\beta,f_r)+V(\beta,\alpha,f_r), \tag{18}$$

$$V(\alpha,f_q,\alpha)+V(\beta,f_q,\beta)\le V(\alpha,f_q,\beta)+V(\beta,f_q,\alpha), \tag{19}$$

$$V(f_p,\alpha,\alpha)+V(f_p,\beta,\beta)\le V(f_p,\alpha,\beta)+V(f_p,\beta,\alpha), \tag{20}$$

For (18), because g(x) increase monotonously when x > 0 , combining Eq. (17), we just need to prove following inequality:

$$\left|\alpha-f_r\right|+\left|\beta-f_r\right|\le\left|\alpha-2\beta+f_r\right|+\left|\beta-2\alpha+f_r\right|. \tag{21}$$

Without loss of generality, supposing $\alpha<\beta$, three possible relationships might exist among α , β and fr, we discuss them respectively as follows. When $\alpha<f_r<\beta$, it yields $|\alpha-f_r|+|\beta-f_r|=\beta-\alpha$, while $|\alpha-2\beta+f_r|+|\beta-2\alpha+f_r|=3(\beta-\alpha)$. In this case, (18) is proper. When $fr<\alpha<\beta$, it yields $|\alpha-f_r|+|\beta-f_r|-|\alpha-2\beta+f_r|=-\beta+2\alpha-fr\le|\beta-2\alpha+f_r|$. So (18) is proper too. While $\alpha<\beta<f_r$, likewise the same conclusion can be acquired.

To prove (19) analogously, only following inequality should be fulfilled:

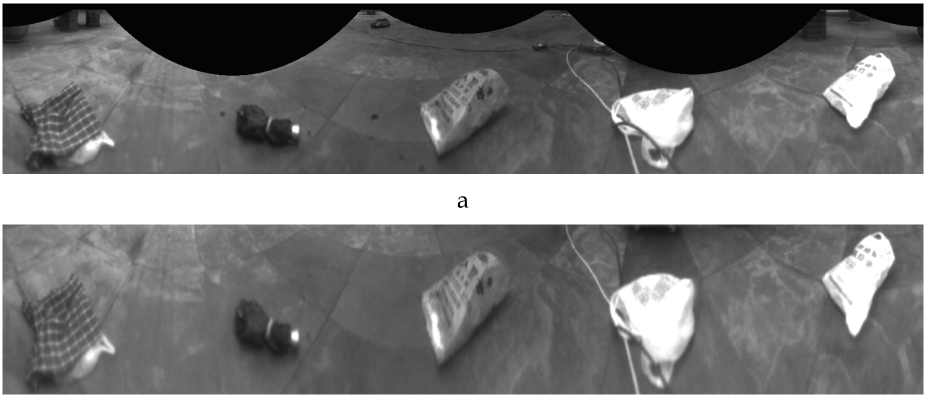
$$2\left|\alpha-f_q\right|+2\left|\beta-f_q\right|\le2\left|\alpha+\beta-2f_q\right|+2\left|\alpha-\beta\right|. \tag{22}$$

Note that the inequality below is invariably true, $|x|+|y|\le|x-y|+|x+y|\le2|x|+2|y|$. Thus (19) can be proved simply. Also we can prove (20) similarly, as (20) is very similar to (19). Now we have proved that our three-variable smoothness model is graph-representable definitely. We use swap algorithm in (Y. Boykov et al., 2001) analogously to optimize our energy model, since the integrated illustration about graph construction has been described in (V. Kolmogorov & R. Zabih, 2004), we only need to specify three possible cases in energy function Vpqr as follows. One case is that $f_p,f_q,f_r\in\{\alpha,\beta\}$, in this case, Vpqr is the typical regular function of three binary variables. The second case is only two of fp, fq, fr belong to $\{\alpha,\beta\}$, in this case, Vpqr is virtually a regular function of two binary variables. The last case is the simplest one in which only one of these three continuous labels belongs to $\{\alpha,\beta\}$, Vpqr is a regular function of one binary variable. Note that above three cases nearly invariably exist in fact. We no longer describe the graph construction process since all the cases are amply described in (V. Kolmogorov & R. Zabih, 2004), although our graph construction process is slightly different when considering the actual model.

4. Stereo Matching

In this section, combining the improved model and corresponding graph cuts algorithm, we present detailed steps for omnidirectional stereo matching.

4.1 Handling Omnidirectional Images



a

b

Figure 3. Unwrapped cylindrical images, of which a corresponding to outer circle image and b inner circle image

The images acquired by our OSVOD (Fig. 2) have some particularities in contrast to normal stereo pairs as follows, which may lead to poor results using traditional stereo matching methods: (1) The upper mirror and nether mirror have different focal length that the camera focal length has to compromise with the two, thus causing defocusing effect. As a result, similarity measures, such as SSD, take on much less discriminability. (2) In this close-quarter imaging, the object surface is always not frontal-parallel to the virtual retina of the camera, resulting in large foreshortening effect between the outer circle image and the inner circle image. (3) Quite a number of weak textured and textureless areas exist in our real indoor scene, more difficulties are bound to bring on in stereo matching. We choose this scene with an eye to the actual ground circumstance, which is usually apt to be weak textured and textureless. (4) The wide-baseline vision system can enhance the calculative precision, whereas, the pending disparity space is larger correspondingly, this increases the search range and ambiguous results tend to bring on. To solve these problems, our method consists of the following several steps: we first convert the raw image both to two cylindrical images and planform images corresponding to images via nether and upper mirrors respectively (Fig. 5 and Fig. 6). The vertical lines with the same abscissa in the cylindrical images are the same epipolar. We compute a similarity measurement for each disparity on every epipolar curve in the cylindrical images. The similarity measurement of a pixel pair is set as the average cost value of that computed from cylindrical images and that from planform images. This is to solve problem (2), as surface perpendicular to the ground tend to have good similarity measurement on the cylindrical images and surface parallel to the ground on the planform images. Second, we choose an appropriate similarity measure described below and also make necessary modification in the iteration of graph cuts to handle panoramic images, including deformed matching template, adaptable template scale. Finally we use improved 3-variable smoothness model via graph cuts to enhance the performance of our algorithm largely. These two steps are to solve problems (1), (3) and (4).

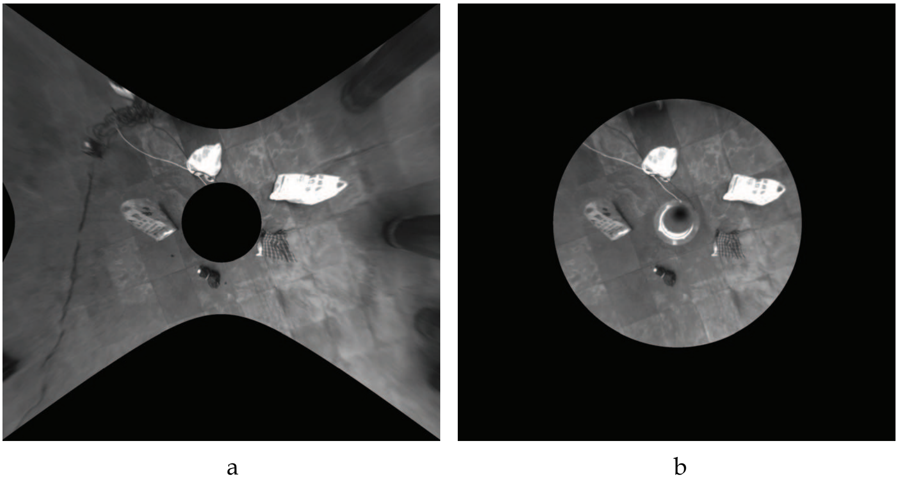


Figure 4. Converted planform images, of which a corresponding to outer circle image and b inner circle image

4.2 MZNCC

The similarity measure we choose here is zero-mean normalized cross correlation (ZNCC), since it is invariant to intensity and contrast between two images. But directly using this measure would result in low discriminability. Chances exist that two templates with great difference in average gray-level or standard deviation which cannot be deemed as matched pair may have high ZNCC value. To avoid this possibility, we modified ZNCC (called MZNCC) by multiplying a window function as follows:

$$MZNCC(p,d)=\frac{\sum(I_a(i,j+d)-\mu_a)\cdot(I_b(i,j)-\mu_b)}{\sigma_a\cdot\sigma_b}\cdot w(|\mu_a-\mu_b|)\cdot w(\frac{\max(\sigma_a,\sigma_b)}{\min(\sigma_a,\sigma_b)}-1) \qquad (23)$$

where $w(x)=\begin{cases}1, & x<\lambda \\ \lambda/x & x\geq\lambda\end{cases}$, μ_a and μ_b are the average grey-level of matching window, σ_a and σ_b are the standard deviation, d denotes the disparity of pixel p.

We define our texure level of each point following the notion of bandwidth of the bandpass filter. For a given pixel and a given template centred in the pixel, we slide the template one pixel at a time in the two opposite directions along the epipolar line and stop at the location the MZNCC value of the shifted template with the primary one decrease below a certain threshold for the first time. Let *l* be the distance between the two stop points, which is inverse proportional the texture level. The definition of texture intensity can be formalized as:

$$Tex(u,v)=\sum_{-r\le(i,j)\le r} {(I(u+i,v+j)-\bar{I})^2}\Big/{l^2} \tag{24}$$

Where r is the radius of the template. With the use of this defined texture intensity and two thresholds, the whole image can be divided into three regions: strong textured, weak textured and textureless regions. Unlike others straightforwardly use sum of intensity difference, we define the data term in our energy function in the form of MZNCC value multiplies a penalty coefficient Cp aim to assign different weights to different points based on the texture level. Generally, as the less reliability of the weak textured area and textureless one, we give following form of Cp.

$$C_p=\begin{cases}\mu_s, & Tex>t_s\\ \mu_w, & t_w\le Tex\le t_s.\\ \mu_l, & t_l<Tex<t_w\end{cases} \tag{25}$$

where $\mu_s>\mu_w>\mu_l$ and t_s,t_w,t_l represent corresponding thresholds to differentiate above three typical areas.

4.3 Template Rectification and Adaptive Scale

For certain corresponding pixel pairs, it is expected that the MZNCC value of the two templates centered at these two points are very close to 1. This expectation is well satisfied when the two image templates are the projections of a single surfaces and this surface is frontal-parallel to the imaging plane of the virtual camera. When larger image templates straddle depth discontinuities, possibly including occluded regions, the MZNCC value may decrease to a value much smaller than 1. Also, if the surface is not parallel to the imaging plane, especially as the ground plane in our scene perpendicular to the imaging plane, the foreshortening effect makes the two templates differ quite significantly, also reduce the MZNCC value to an unsatisfactory amount.

In this iterative framework of graph cuts, it is natural to estimate the appropriate template scale not to straddle depth discontinuities and rectify the template to compensate the foreshortening effect from current temporal result at each step. At each pixel in the image, we first use the largest template scale. We then compute the variance of the depth data in the template. If the variance exceeds an appropriately chosen threshold, it may be that the template scale is too large. Otherwise we continue to make use this template scale for MZNCC calculation.

After the scale is determined, it is ensured that the template corresponds to a single surface. We use the depth data to fit this surface to a plane in 3-D space, and then reproject this plane to the other image. Normally, the reprojected template is not a rectangle any more if the surface is not frontal-parallel. And we compute the MZNCC value between the rectangle template in the reference image and the rectified template in the other image. In this way, foreshortening effect is well compensated.

4.4 Graph Cuts

Considering the high-performance results in stereo matching, eventually, we use improved graph cuts based on three-variable smoothness model to optimize our energy function. Since the critical part of energy function has been discussed in section 3, now we briefly present the general form corresponding to our omnidirectional stereo. We use the appropriate energy function *E* in the form (10) to accommodate close-quarter measurement. $E_{data}(f)$ describes the MZNCC of corresponding pixels, and we also use different penalty coefficients to distinguish pixels with different texture.

$$E_{data}(f)=\sum(1-C_p \cdot MZNCC(p,f_p)), \tag{26}$$

E_{smooth} introduces a penalty for neighboring pixels having different disparity values.

$$E_{smooth}(f)=\sum_{(p,q,r)\in N} C_{pqr} \cdot \min(T,g(\left|f_p+f_r-2f_q\right|+\left|f_p-f_r\right|)). \tag{27}$$

Note that we do not take occlusion into account when considering no occlusions come forth since we invariably choose the unwrapped cylindrical image, which corresponds to the outer image reprojected by nether mirror, as the reference image in stereo. You see, as the upper mirror is higher than the nether one, any point that can be seen in outer image must be seen in inner image unless it exceeds the view area in the camera or the theoretical matching point can not exist. In this case, we call it falls in 'blind area', virtually this always comes true in our omnidirectional images and we invariably neglect the pixels in 'blind area'.

5. Experimental Results

In this section, we present our experimental results. In data term E_{data} ,Cp can be defined as follows:

$$C_p=\begin{cases} 1, & Tex>10 \\ 0.5 & 2\le Tex\le 10, \\ 0.25 & 0<Tex<2 \end{cases}$$

in view of E_{smooth} , Cpqr =0.01, T=100, g(x)=x2 .

To observe the experimental results legibly, we utilize a part of the real images (Fig. 7(a), 200*200) in Fig. 3 to test our algorithm. Fig.8 shows the experimental results via two typical 2-variable smoothness models and our 3-variable one. Fig. 7(b) gives the initial depth map by MZNCC via winner-take-all (WTA). Fig. 7(c) shows the depth map based on the Potts model. Fig. 7(d) denotes the results of traditional 2-variablated convex priors. Fig. 7(e) and Fig. 7(f) present our results based on 3-variable truncated convex priors model respectively. The 2-variable model, which results in smaller gaps in depth map, performs better than Potts model. Notice that our algorithm produces an answer which varies most smoothly and consistently in three models because our energy function can approximate the true scene better, especially in our omnidirectional images which possess some particularities we have mentioned above. Note that when introducing template rectification and adapt scale (f) performances better at the boundary than (e), however, due to the imprecise initial depth map, the amelioration is not very obvious but unspent. Unlike some standard images, obtaining precise groundtruth depth map for our real images is quite complicated, therefore, it is hard to describe our results quantitatively. Note that in all the experimental results, we neglect the 'blind area', in which we simply set a certain depth value.

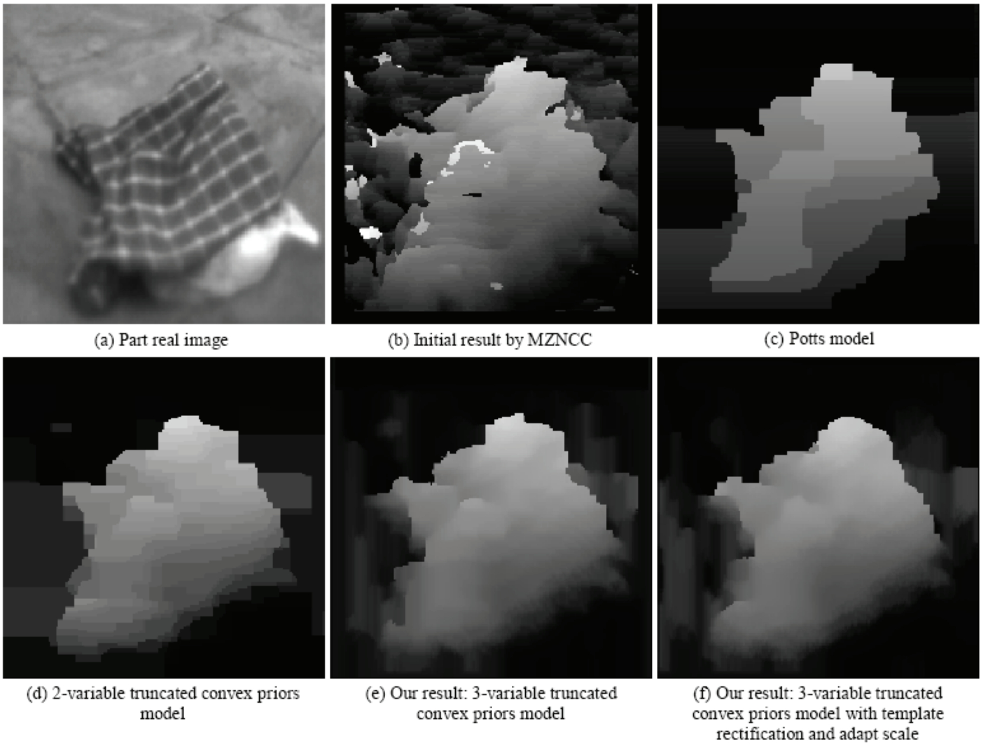

(a) Part real image
(b) Initial result by MZNCC
(c) Potts model
(d) 2-variable truncated convex priors model
(e) Our result: 3-variable truncated convex priors model
(f) Our result: 3-variable truncated convex priors model with template rectification and adapt scale

Figure 7. Depth maps of part omnidirectional images

Finally, we give the complete depth map in Fig. 8 and also the corresponding obstacle sketch map consisting of point clouds can be found in Fig. 9. Fig. 8 describes the stereo correspondence of images in Fig. 7 precisely and this preferable performance is reliable for depth map regeneration, even though the obstacle we placed on ground is somewhat low, which results in relatively difficulties in the reconstructed process. Fig. 9 shows an effective obstacle sketch map for a mobile robot. These point clouds denote existence of obstacle, we can see the resolution gets lower when the object is far from the center (maybe need to zoom in to see clearly), this was also mentioned in (G. Jang, S. Kim & I. Kweon, 2005). To get a reasonable obstacle sketch map, we necessarily set a small threshold (about 10mm in this paper) to judge if the corresponding point should be seen as an obstacle point since the error nearly invariably exists.

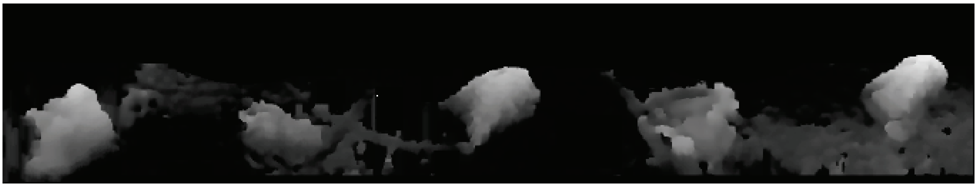


Figure 8. Final depth map (Fig.5(a) is reference image)

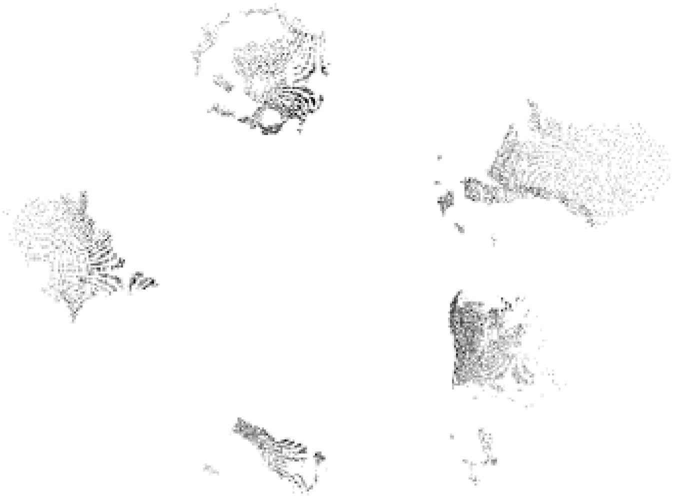


Figure 9. Final omnidirectional obstacle sketch map

6. Summary

In this paper, we have developed a graph-representable three-variable smoothness model for graph cuts to fit the smooth assumption for our omnidirectional images taken by a novel vision sensor. We further develop MZNCC as a suitable similarity measurement and also the necessary modification, including deformed matching template and adaptable scale. Experiments demonstrate the effectiveness of our algorithm, based on which, the regenerated obstacle map is finer for a mobile robot.

7. References

A. Jagmohan, M. Singh and N. Ahuja(2004). Dense Two View Stereo Matching Using Kernel Maximum Likelihood Estimation, *Proc. of ICPR'04*. pages 28-31, 2004.

D. Scharstein and R. Szeliski(2002). A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*. 2002

Devernay and F. Faugeras(1994). Computing Differential Properties of 3-D Shapes from Stereoscopic Images without 3-D Models. Proc. of CVPR'94. pages 208–213, 1994.

Fabrizio, J.; Tarel, J. & Benosman, R. (2002). Calibration of Panoramic Catadioptric Sensors Made Easier, *Proceedings of Workshop on Omnidirectional Vision*, pp. 45-52, Copenhagen, Denmark, Jun 2002

Geyer, C. & Daniilidis, K.(1999). Catadioptric Camera Calibration, *Proceedings of International Conference on Computer Vision*, Vol. 1, pp. 398-404, Kerkyra, Greece, Sep 1999

Geyer, C. & Daniilidis, K. (2000). Paracatadiopric Camera Calibration, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 5, pp. 687-695

Geyer, C. & Daniilidis, K. (2003). Conformal Rectification of Omnidirectional Stereo Pairs, *Proceedings of IEEE Workshop on Omnidirectional Vision and Camera Networks*, pp. St. Louis, USA, Jun 2003

G. Jang, S. Kim and I. Kweon(2005). Single Camera Catadioptric Stereo System. *Proc. Of Workshop on Omnidirectional Vision, Camera Networks and Nonclassical cameras(OMNIVIS2005),* 2005

H. Ishikawa(2003). Exact optimization for markov random fields with convex priors. *IEEE Trans. on PAMI*. 25(10):1333–1336, October 2003.

H. Koyasu, J. Miura and Y. Shirai(2002). Recognizing Moving Obstacles for Robot Navigation Using Real-time Omnidirectional Stereo Vision. *Journal of Robotics and Mechatronics*, 14 (2), pages 147-156, 2002.

J. Gluckman, S. K. Nayar and K. J. Thoresz(1998). Real-time Omnidirectional and Panoramic Stereo. Proc. of DARPA Image Understanding Workshop. pages 299-303, 1998.

E. L. L. Cabral, J. C. de Souza Junior and M. C. Hunold. Omnidirectional Stereo Vision with a Hyperbolic Double Lobed Mirror, Proc. of ICPR'04, 2004.

Kang, S.B. (2000). Catadioptric Self-calibration, *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 201-207, Hilton Head Island, South Carolina, Jun 2000

K. Nayar(2004). Catadioptric Omnidirectional Camera. *Proc. Of CVPR'97*, pages 482-488, 1997.

K. Tan, H. Hua and N. Ahuja(2004). Multiview Panoramic Cameras Using Mirror Pyramids, IEEE Trans. on PAMI. 26(6), 2004.

Luong, Q.T. & Faugeras, O.D. (1996). The Fundamental Matrix: Theory, Algorithms, and Stability Analysis. *Int. J. of Computer Vision*, Vol. 17, No. 1, pp. 43-76

M.F.D. Southwell, A. Basu and J. Reyda. Panoramic Stereo. *Proc. of ICPR'96*, pages 378-382, 1996.

Miousik, B. & Pajdla, T. (2003a). Estimation of Omnidirectional Camera Model from Epipolar Geometry, *Proceedings of 2003 IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 485-490, Madison, USA, Jun 2003

Miousik, B. & Pajdla, T. (2003b). Omnidirectional Camera Model and Epipolar Geometry Estimation by RANSAC with Bucketing, *Proceedings of Scandinavian Conf. on Image Analysis*, pp. 83-90, Goteborg, Sweden, Jun 2003

Pedro F. Felzenszwalb and Daniel P. Huttenlocher(2006). *International Journal of Computer Vision*, 70(1), 2006.

R. Bunschoten and B. Kr¨ose(2003). Robust scene reconstruction from an omnidirectional vision system. *IEEE Transactions on Robotics and Automation.* pages 351– 357, 2003.

Sara, R(2002). Finding the Largest Unambiguous Component of Stereo Matching. *Proc. of ECCV'02*, pages 900–914, 2002.

Sooyeong Yi and Narendra Ahuja(2006). An Omnidirectional Stereo System Using a Single Camera. *Proc. of ICPR ’06*, 2006.

Su, L. & Zhu, F. (2005). Design of a Novel Stereo Vision Navigation System for Mobile Robots, *Proceedings of IEEE Conference on Robotics and Biomimetics*, No. 1, pp. 611-614, Hong Kong, Jun 2005

Sven Fleck, Florian Busch, Peter Biber, Wolfgang Straßer and Henrik Andreasson(2005). Omnidirectional 3D Modeling on a Mobile Robot using Graph Cuts. *Proc. of ICRA ’05*. 2005.

T.L. Conroy and J.B. Moore(1999). Resolution Invariant Surfaces for Panoramic Vision Systems. *Proc. of ICCV ’99*, pages 392-397, 1999.

V. Kolmogorov and R. Zabih(2001). Computing visual correspondence with occlusions using graph cuts. *Proc. Of ICCV ’01*. pages 508-515, 2001.

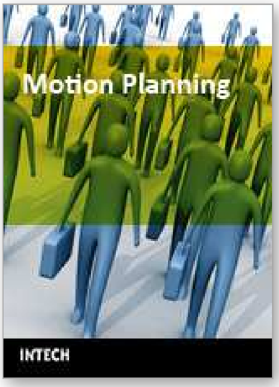
V. Kolmogorov and R. Zabih(2004). What Energy Functions Can Be Minimized via Graph Cuts? *IEEE Trans. on PAMI*. 26(2), 2004.

Y. Boykov, O. Veksler and R. Zabih(2001). Fast approximate energy minimization via graph cuts. *IEEE Trans. on PAMI*. pages 1222–1239, 2001.

Yedidia, J.S., Freeman, W.T., Weiss, Y.(2000). Generalized belief propagation. *Proc. of NIPS ’00*. pages 689-695, 2000.

Ying, X. & Hu, Z. (2004). Catadioptric Camera Calibration Using Geometric Invariants, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 10, pp. 1260-1271

Zhang, Z. & Faugeras, O. (1997). An Effective Technique for Calibrating a Bincular Stereo Through Projective Reconstruction Using Both a Calibration Object and the Environment, *Journal of Computer Vision Research*, Vol. 1, No. 1, pp. 58-68



Motion Planning
Edited by Xing-Jian Jing

ISBN 978-953-7619-01-5
Hard cover, 598 pages
Publisher InTech
Published online 01, June, 2008
Published in print edition June, 2008

In this book, new results or developments from different research backgrounds and application fields are put together to provide a wide and useful viewpoint on these headed research problems mentioned above, focused on the motion planning problem of mobile ro-bots. These results cover a large range of the problems that are frequently encountered in the motion planning of mobile robots both in theoretical methods and practical applications including obstacle avoidance methods, navigation and localization techniques, environmental modelling or map building methods, and vision signal processing etc. Different methods such as potential fields, reactive behaviours, neural-fuzzy based methods, motion control methods and so on are studied. Through this book and its references, the reader will definitely be able to get a thorough overview on the current research results for this specific topic in robotics. The book is intended for the readers who are interested and active in the field of robotics and especially for those who want to study and develop their own methods in motion/path planning or control for an intelligent robotic system.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Lei He, Chuanjiang Luo, Feng Zhu and Yingming Hao (2008). Stereo Matching and 3D Reconstruction via an Omnidirectional Stereo Sensor, Motion Planning, Xing-Jian Jing (Ed.), ISBN: 978-953-7619-01-5, InTech, Available from: http://www.intechopen.com/books/motion_planning/stereo_matching_and_3d_reconstruction_via_an_omnidirectional_stereo_sensor

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821